Inapproximability of the Smallest Superpolyomino Problem*

Andrew Winslow†
awinslow@cs.tufts.edu

1 Introduction

We consider the *smallest superpolyomino problem*: given a set of colored polyominoes, find the smallest superpolyomino containing each input polyomino as a subpolyomino. Alternatively, find an overlapping arrangement of the polyominoes such that all overlapping cells have matching colors and the union of the polyominoes is as small as possible.

In one dimension, this problem is equivalent to the *smallest superstring problem* and admits a greedy constant-factor approximation algorithm [1]. Charikar et al. [2] use this to develop a straightforward $O(\log^3 n)$-approximation algorithm for finding the smallest context-free grammar encoding a string.

One motivation for investigating the smallest superpolyomino problem is the possibility of extending the Charikar algorithm to higher dimensions, yielding good grammar-based image and shape compression algorithms. Here we show that such an extension is unlikely to exist by proving that the smallest superpolyomino problem is NP-hard to approximate within a $O(n^{1/3-\varepsilon})$-factor for any $\varepsilon > 0$ by a reduction from chromatic number.

2 Definitions

A polyomino $P = (S, L)$ is defined by a connected set of points S on the square lattice (called *cells*) containing $(0, 0)$, and a coloring of the cells, e.g. cell $(3, 1)$ is red, cell $(3, 2)$ is gray, etc. We denote the color of the cell (x, y) as $P(x, y)$, and $|P|$ denotes the number of cells in P, i.e. the *size* of P. Two polyominoes $P_u = (S_u, L_u)$ and $P_v = (S_v, L_v)$ at some translation (δ_x, δ_y) are *compatible* if for each (x, y), either $P_v(x, y)$ or $P_u(x, y)$ is empty or $P_v(x, y) = P_u(x+\delta_x, y+\delta_y)$. Similarly, a polyomino $P = (S, L)$ is a *superpolyomino* of $P' = (S', L')$ if there exists a translation (δ_x, δ_y) such that for each (x, y), either $(x, y) \notin S'$ or $P'(x + \delta_x, y + \delta_y) = P(x, y)$, i.e. there is a translation of P' such that P' is compatible with P and lies entirely in P.

3 Reduction

Given a graph $G = (V, E)$, each vertex $v \in V$ is converted into a polyomino $P_v = (S_v, L_v)$ that encodes v and the neighbors of v in G (see Figure 1). Each P_v is a rectangular $2|V| \times |V|$ polyomino with up to $|V| - 1$ single squares removed and lower-left corner at $(0, 0)$. The four corners of all P_v have a common set of four colors: green, blue, purple, and orange. Cells at locations $\{(2i + 1, 1) \mid 0 \le i < |V|\}$ are colored black if $v_i = v$, red if $(v, v_i) \in E$, or are empty locations if v_i is not v or a neighbor of v. All remaining cells have a common gray color.

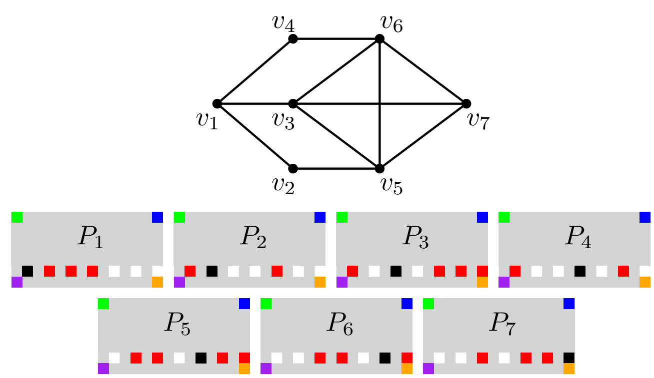

Figure 1: An example of the set of polyominoes generated from an input graph by the reduction.

Consider how two polyominoes P_u and P_v can overlap, depending upon the relationship of u and v. Because of the four distinct corner colors, P_u and P_v can only overlap when these four locations in P_v are translated to the same locations in P_u. In this translation, the cells at location $(2i + 1, 1)$ in P_u and P_v are compatible exactly when $(u, v) \notin E$, i.e. u and v are not neighbors. All other cells are colored gray and thus compatible.

The superpolyomino formed by a pair of compatible P_u and P_v in this translation has the common set of four colored corner cells and many gray cells, and has two black cells and a number of red cells corresponding to the combined neighborhoods of u and v. Then by induction, any set of polyominoes can overlap if and only if they form an independent set. Moreover, if they overlap, they overlap using a set of translations in which

*A full version of this paper is available at http://arxiv.org/abs/1210.3877.

†Department of Computer Science, Tufts University. Research supported in part by NSF grants CCF-0830734 and CBET-0941538.

the four corners of all polyominoes are placed at four common locations.

Because the polyominoes can only overlap in this constrained way, any superpolyomino of the polyominoes $\{P_v \mid v \in V\}$ consists of a number of *decks* of superimposed polyominoes corresponding to independent sets of vertices in G arranged disjointly to form a single connected polyomino (see Figure 2).

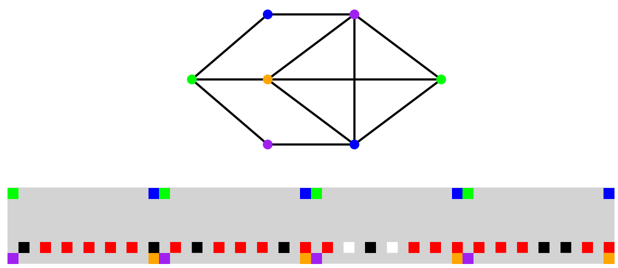

Figure 2: An example of a corresponding 4-deck superpolyomino and 4-colored graph.

Recall that each P_v is a $2|V| \times |V|$ rectangle with $|V|$ cells colored black, red, or are not present. The size of P_v is then between $2|V|^2 - |V| + 1$ and $2|V|^2$ depending upon the number of neighbors of v, and each deck of polyominoes also has size in this range.

Lemma 3.1 *For a graph $G = (V, E)$, there exists a superpolyomino of size at most $2k|V|^2$ for polyominoes $\{P_v \mid v \in V\}$ if and only if the vertices of V can be k-colored.*

Proof First, consider extreme sizes of superpolyominoes consisting of k and $k-1$ decks. For any V and k with $1 \leq k \leq |V|$, $(k-1)(2|V|^2) = 2k|V|^2 - 2|V|^2 < 2k|V|^2 - k|V| = k(2|V|^2 - |V|)$, i.e. the size of any superpolyomino of $k-1$ decks is smaller than the size of any superpolyomino of k decks.

We now prove both implications of the lemma. First, assume the superpolyomino of size at most $2k|V|^2$ exists. Then the superpolyomino must consist of at most k decks. Each deck is the superposition of a set of polyominoes forming an independent set, so G can be k-colored.

Next, assume that G can be k-colored. Then the polyominoes $\{P_v \mid v \in V\}$ can be translated to form k decks, one for each color, each with size at most $2|V|^2$. Placing these decks adjacent to each other yields a superpolyomino of size at most $2k|V|^2$. □

Note that only $|V|$ cells of each P_v are distinct and depend on v, while the other $2|V|^2 - |V|$ are held constant. The extra cells are needed for the first inequality in Lemma 3.1, and they effectively "drown out" the difference in sizes of various decks due to the number of cells not present in each deck.

Theorem 3.2 *The smallest superpolyomino problem is* NP*-hard to approximate within a factor of $O(n^{1/3-\varepsilon})$ for any $\varepsilon > 0$.*

Proof Consider the smallest superpolyomino problem for the polyominoes generated from a graph $G = (V, E)$ with chromatic number k. There are $|V|$ of these polyominoes, each of size $\Theta(|V|^2)$, so the polyominoes have total size $n = \Theta(|V|^3)$. By Lemma 3.1, a superpolyomino of size between $(2|V|^2 - |V|)k'$ and $2|V|^2k'$ exists if and only if there exists a k'-coloring of G. Then by Zuckerman [3], finding a superpolyomino such that $(2|V|^2 - |V|)k'/(2|V|^2)k = O(|V|^{1-\varepsilon}) = \Theta(n^{1/3-\varepsilon})$ is NP-hard.

As seen in Figure 3, the result also holds when constrainted to sets of polyominoes using at most two colors by converting each cell into a unique 8×8 macro-cell.

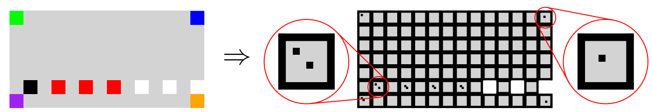

Figure 3: Converting a reduction polyomino (left) to a two-color reduction polyomino (right).

We mention (but do not prove here) that the problem constrained to single-color sets of polyominoes is NP-hard by a reduction from set cover. An example of a polyomino set used in the reduction is seen in Figure 4.

Figure 4: The set of polyominoes produced from the reduction from minimum set cover to smallest superpolyomino for the set $\{\{1, 2\}, \{1, 4\}, \{2, 3, 4\}, \{2, 4\}\}$.

References

[1] A. Blum, T. Jiang, M. Li, J. Tromp, M. Yannakakis, Linear approximation of shortest superstrings, *Journal of the ACM*, 41(4) (1994), 630–647.

[2] M. Charikar, E. Lehman, A. Lehman, D. Liu, R. Panigrahy, M. Prabhakaran, A. Sahai, a. shelat, The smallest grammar problem, *IEEE Transactions on Information Theory*, 51(7) (2005), 2554–2576.

[3] D. Zuckerman, Linear degree extractors and the inapproximability of max clique and chromatic number, *Theory of Computation*, 3 (2007), 103–128.